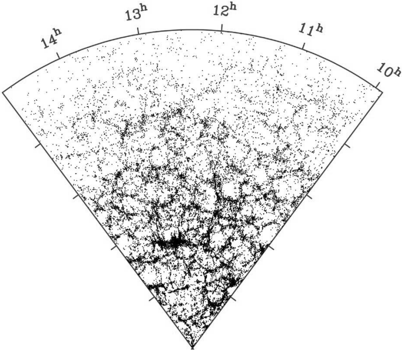
Analyse topologique de données et estimation de support

Eddie Aamari
Laboratoire de Probabilité, Statistiques et Modélisation
CNRS, Université Paris Cité, Sorbonne Université

Congrès français de mécanique
—
Nantes

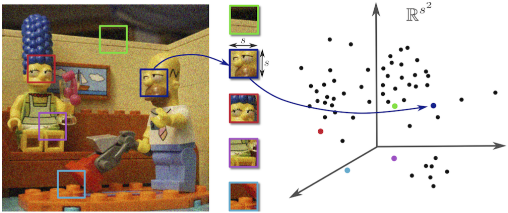
30 août 2022

Data with a Global Geometric Structure



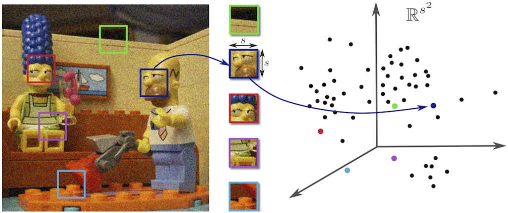
Large Scale Galaxy Structures: one point represents a galaxy in $\mathbb{R}^3$
[2dF Galaxy Redshift Survey]

Data with a Global Geometric Structure

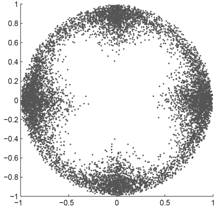


Extracting all the $s \times s$ patches of an image with $m \times n$ pixels.
[Houdard – 2018]

Data with a Global Geometric Structure

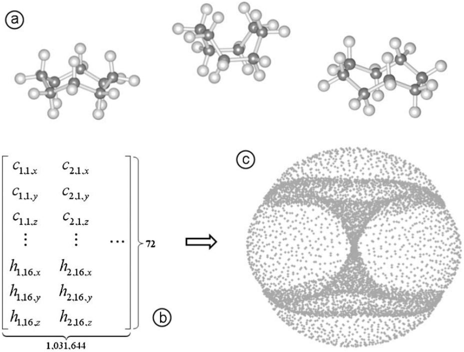


Extracting all the $s \times s$ patches of an image with $m \times n$ pixels.
[Houdard – 2018]



For $s = 7$, one image $M \in \mathbb{R}^{n \times m}$ yields $\asymp mn$ points in $\mathbb{R}^{7 \times 7} = \mathbb{R}^{49}$
[Xia – 2016]

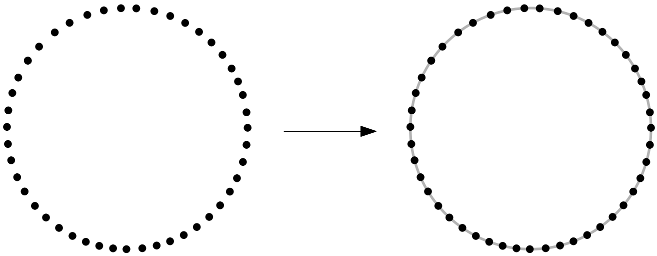
Data with a Global Geometric Structure



Cyclo-octane (C_8H_{16}) conformations
[Martin *et al.* – 2010]

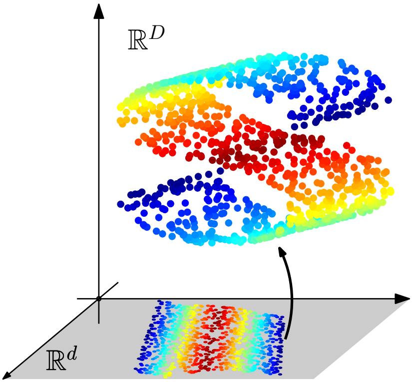
One conformation is described with a point in $\left(\mathbb{R}^3\right)^{8+16} = \mathbb{R}^{72}$.

Uncover Data Structure



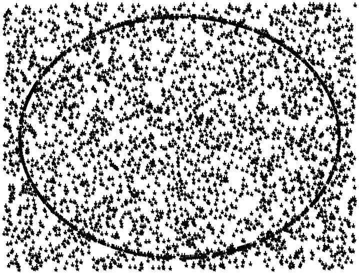
Input: a set $\mathbb{X}_n = \{X_1, \dots, X_n\}$ of observations.
Goal: Understand the underlying structure of the data, for interpretation or summary.

Challenge 1: Dimension



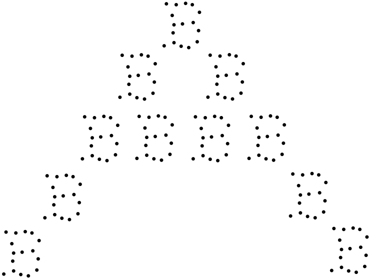
What dimension is this S-shape?

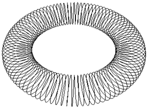
Challenge 2: Noise

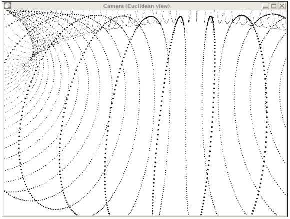


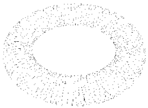
Are my data corrupted?

Challenge 3: Scale





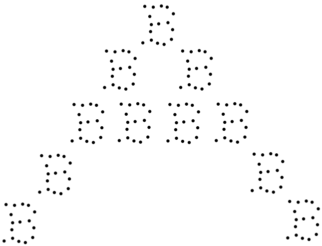


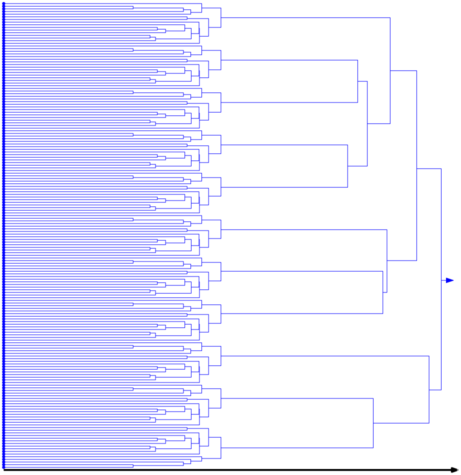


Zoom in or zoom out?

Dendrogram, Persistence Barcodes and Generalization

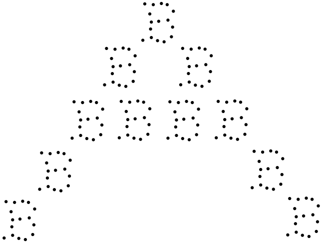
$$\begin{aligned} \mathrm{d}_{\mathcal{P}} : \quad & \mathbb{R}^2 \to \mathbb{R} \\ & x \mapsto \min_{p \in \mathcal{P}} \|x - p\| \end{aligned}$$

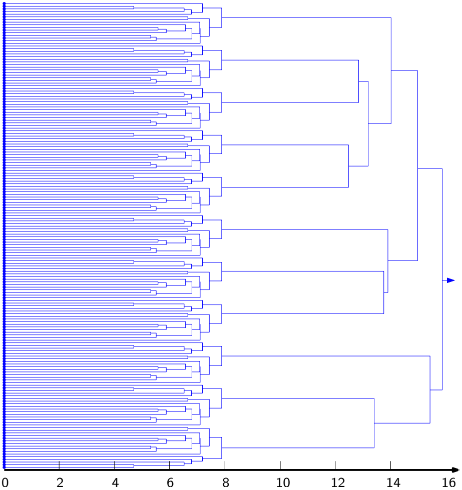




Dendrogram, Persistence Barcodes and Generalization

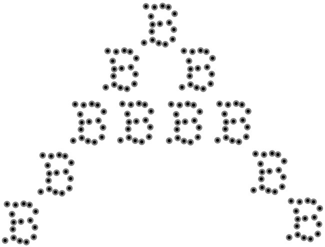
$$\mathrm{d}_{\mathcal{P}} : \quad \mathbb{R}^2 \to \mathbb{R}$$
$$x \mapsto \min_{p \in \mathcal{P}} \|x - p\|$$

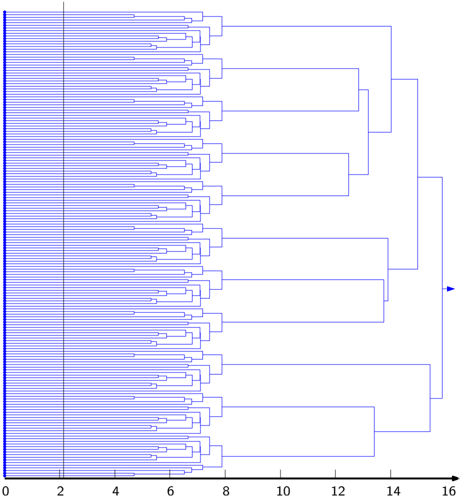




Dendrogram, Persistence Barcodes and Generalization

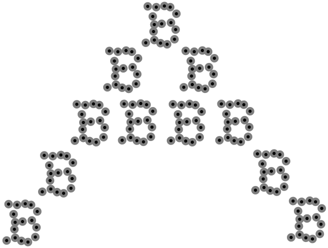
$$\begin{aligned} d_{\mathcal{P}} : \quad & \mathbb{R}^2 \to \mathbb{R} \\ & x \mapsto \min_{p \in \mathcal{P}} \|x - p\| \end{aligned}$$

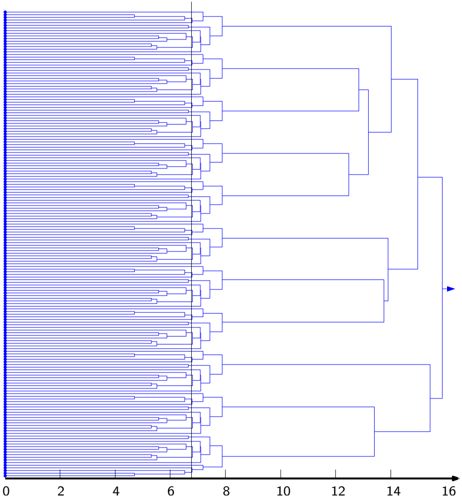




Dendrogram, Persistence Barcodes and Generalization

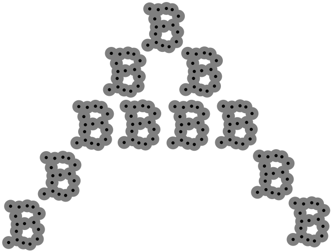
$$\mathrm{d}_{\mathcal{P}} : \quad \mathbb{R}^2 \to \mathbb{R}$$
$$x \mapsto \min_{p \in \mathcal{P}} \|x - p\|$$

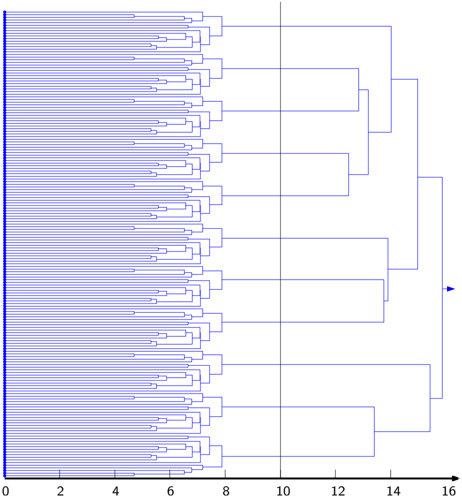




Dendrogram, Persistence Barcodes and Generalization

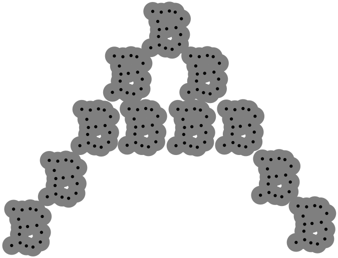
$$\mathrm{d}_{\mathcal{P}}: \quad \mathbb{R}^2 \to \mathbb{R}$$
$$x \mapsto \min_{p \in \mathcal{P}} \|x - p\|$$

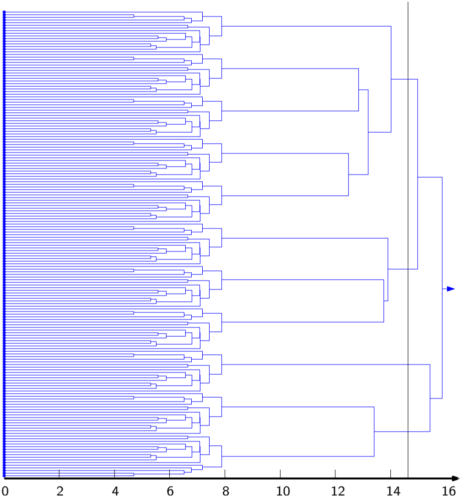




Dendrogram, Persistence Barcodes and Generalization

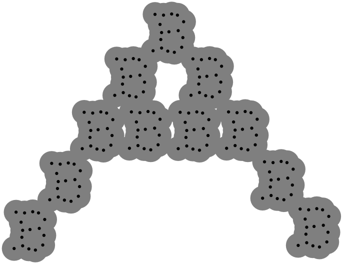
$$\begin{aligned} \mathrm{d}_{\mathcal{P}} : \quad & \mathbb{R}^2 \to \mathbb{R} \\ & x \mapsto \min_{p \in \mathcal{P}} \|x - p\| \end{aligned}$$

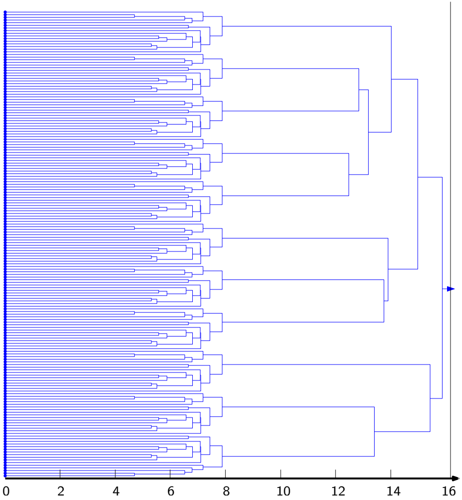




Dendrogram, Persistence Barcodes and Generalization

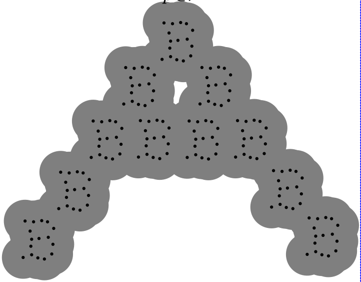
$$\begin{aligned} d_{\mathcal{P}}: \quad & \mathbb{R}^2 \to \mathbb{R} \\ & x \mapsto \min_{p \in \mathcal{P}} \|x - p\| \end{aligned}$$

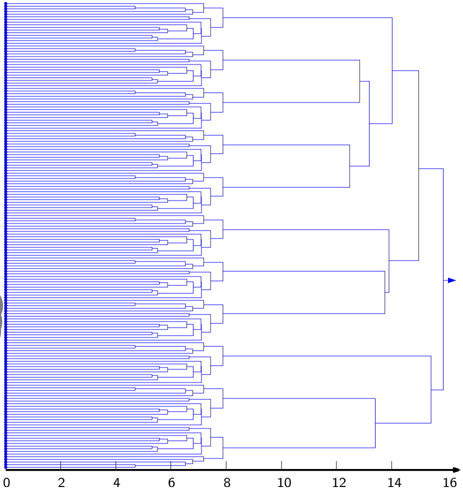




Dendrogram, Persistence Barcodes and Generalization

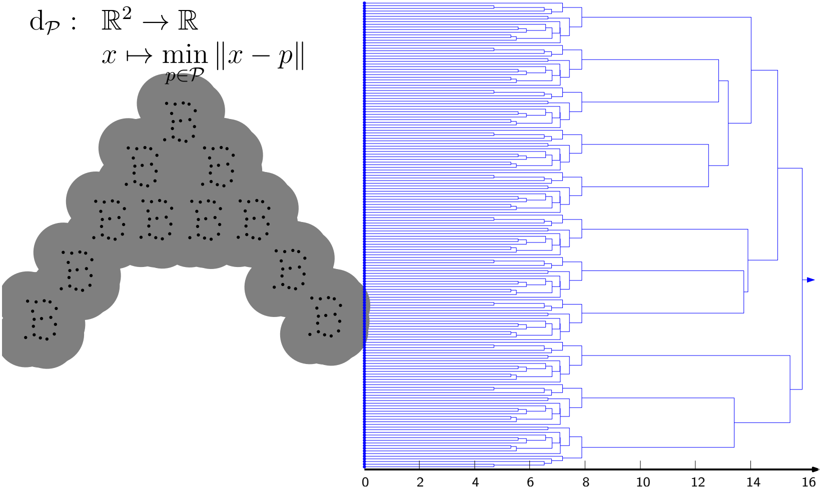
$$\mathrm{d}_{\mathcal{P}} : \quad \mathbb{R}^2 \to \mathbb{R}$$
$$x \mapsto \min_{p \in \mathcal{P}} \|x - p\|$$





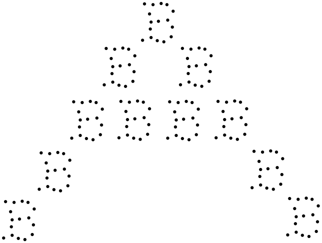
Dendrogram, Persistence Barcodes and Generalization

$$\mathrm{d}_{\mathcal{P}} : \quad \mathbb{R}^2 \to \mathbb{R}$$
$$x \mapsto \min_{p \in \mathcal{P}} \|x - p\|$$



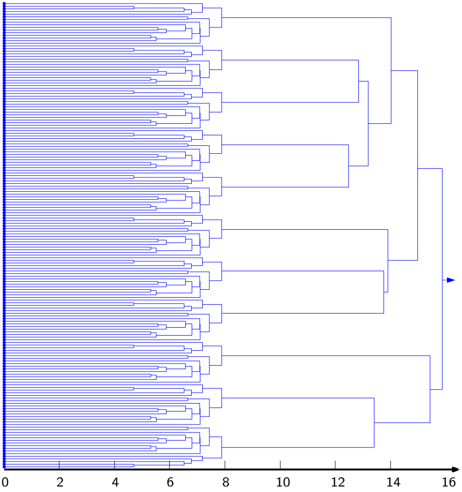
Dendrogram, Persistence Barcodes and Generalization

$$\mathrm{d}_{\mathcal{P}}: \quad \mathbb{R}^2 \to \mathbb{R}$$
$$x \mapsto \min_{p \in \mathcal{P}} \|x - p\|$$



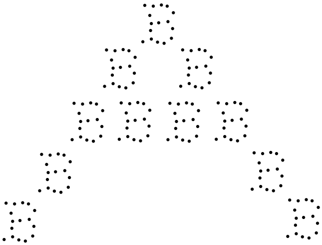
Dendrogram is:

- informative
- unstable

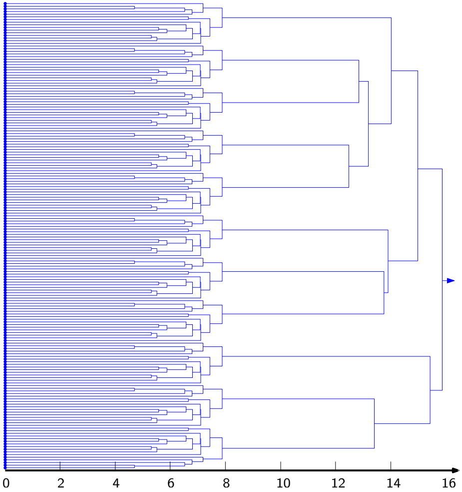


Dendrogram, Persistence Barcodes and Generalization

$$\mathrm{d}_{\mathcal{P}}: \quad \mathbb{R}^2 \to \mathbb{R}$$
$$x \mapsto \min_{p \in \mathcal{P}} \|x - p\|$$

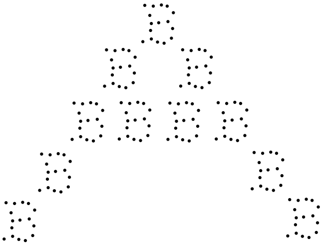


Dendrogram $\to$ barcode

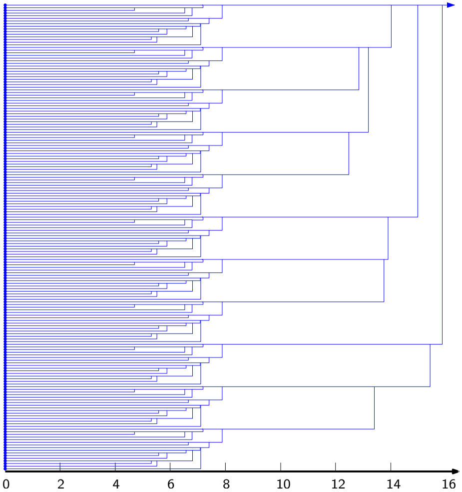


Dendrogram, Persistence Barcodes and Generalization

$$\begin{aligned} \mathrm{d}_{\mathcal{P}}: \quad & \mathbb{R}^2 \to \mathbb{R} \\ & x \mapsto \min_{p \in \mathcal{P}} \|x - p\| \end{aligned}$$

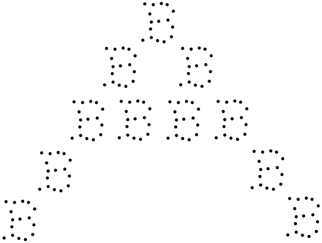


Dendrogram $\to$ barcode

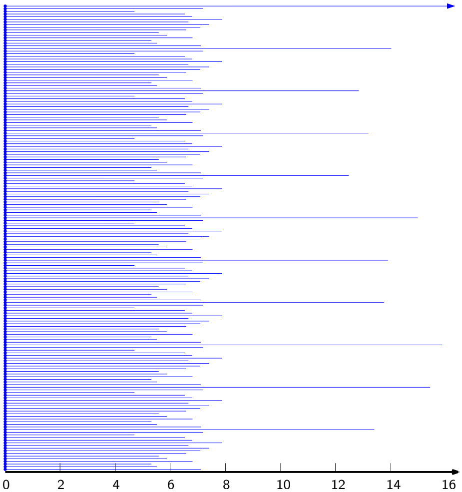


Dendrogram, Persistence Barcodes and Generalization

$$\mathrm{d}_{\mathcal{P}} : \quad \mathbb{R}^2 \to \mathbb{R}$$
$$x \mapsto \min_{p \in \mathcal{P}} \|x - p\|$$

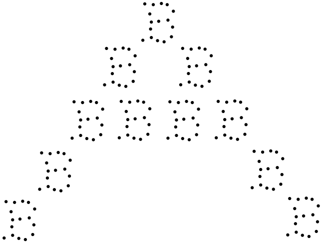


Dendrogram $\rightarrow$ barcode



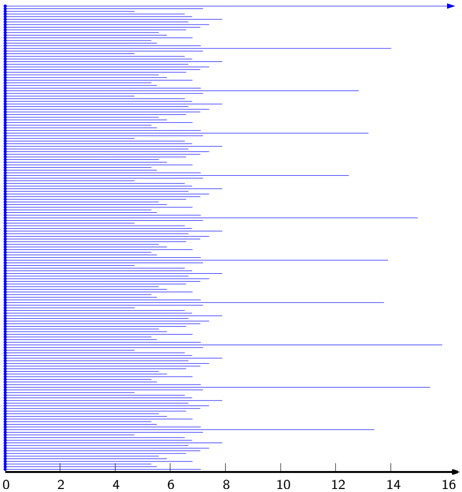
Dendrogram, Persistence Barcodes and Generalization

$$\mathrm{d}_{\mathcal{P}} : \quad \mathbb{R}^2 \to \mathbb{R}$$
$$x \mapsto \min_{p \in \mathcal{P}} \|x - p\|$$



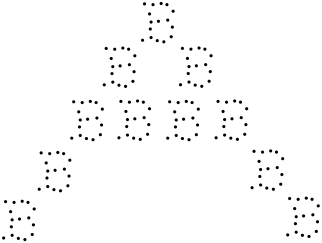
Barcode is:

- less informative
- more stable



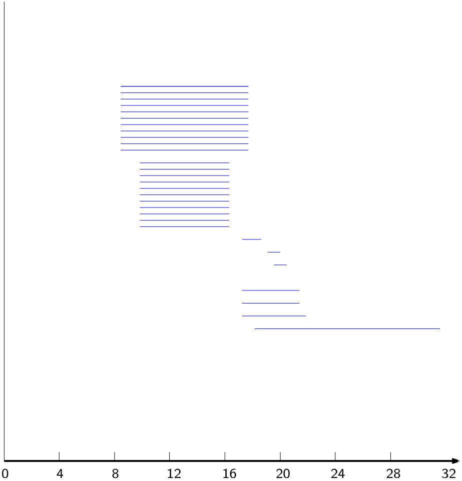
Dendrogram, Persistence Barcodes and Generalization

$$\begin{aligned} \mathrm{d}_{\mathcal{P}} : \quad & \mathbb{R}^2 \to \mathbb{R} \\ & x \mapsto \min_{p \in \mathcal{P}} \|x - p\| \end{aligned}$$



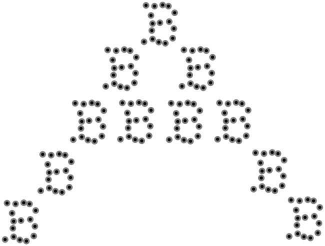
Barcode is:

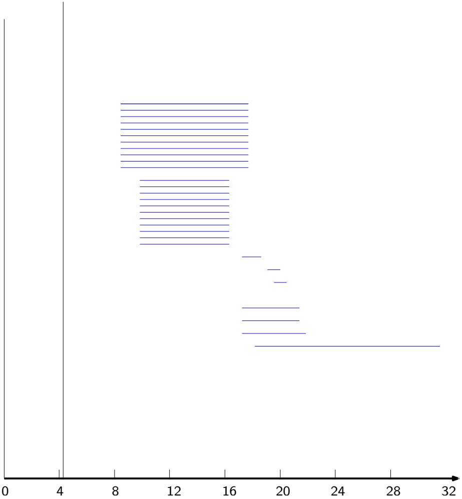
- less informative
- more stable
- generalizable



Dendrogram, Persistence Barcodes and Generalization

$$\begin{aligned} d_{\mathcal{P}} : \quad & \mathbb{R}^2 \to \mathbb{R} \\ & x \mapsto \min_{p \in \mathcal{P}} \|x - p\| \end{aligned}$$



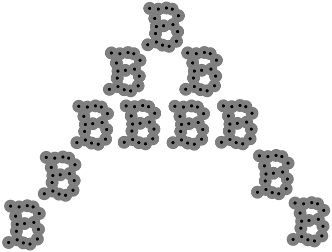


Barcode is:

- less informative
- more stable
- generalizable

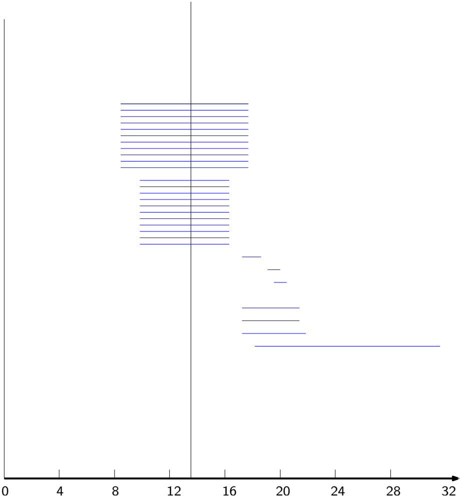
Dendrogram, Persistence Barcodes and Generalization

$$\mathrm{d}_{\mathcal{P}} : \quad \mathbb{R}^2 \to \mathbb{R}$$
$$x \mapsto \min_{p \in \mathcal{P}} \|x - p\|$$



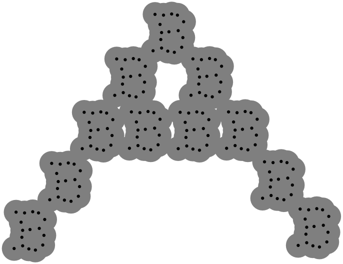
Barcode is:

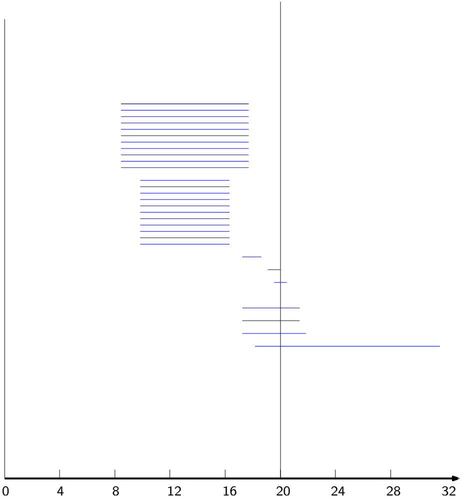
- less informative
- more stable
- generalizable



Dendrogram, Persistence Barcodes and Generalization

$$\begin{aligned} \mathrm{d}_{\mathcal{P}} : \quad & \mathbb{R}^2 \to \mathbb{R} \\ & x \mapsto \min_{p \in \mathcal{P}} \|x - p\| \end{aligned}$$



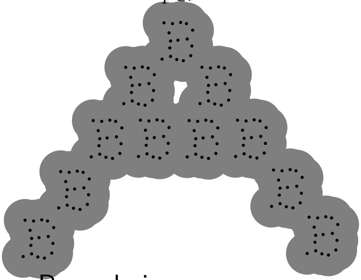


Barcode is:

- less informative
- more stable
- generalizable

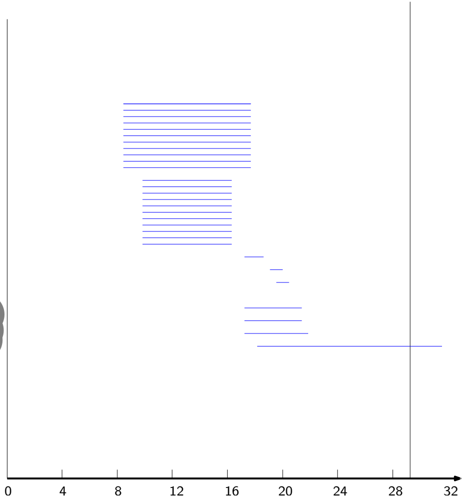
Dendrogram, Persistence Barcodes and Generalization

$$\mathrm{d}_{\mathcal{P}}: \quad \mathbb{R}^2 \to \mathbb{R}$$
$$x \mapsto \min_{p \in \mathcal{P}} \|x - p\|$$



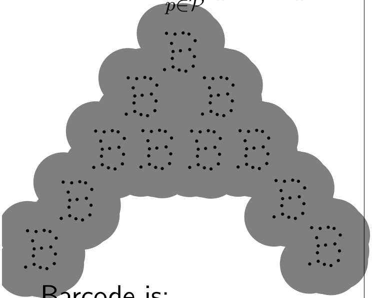
Barcode is:

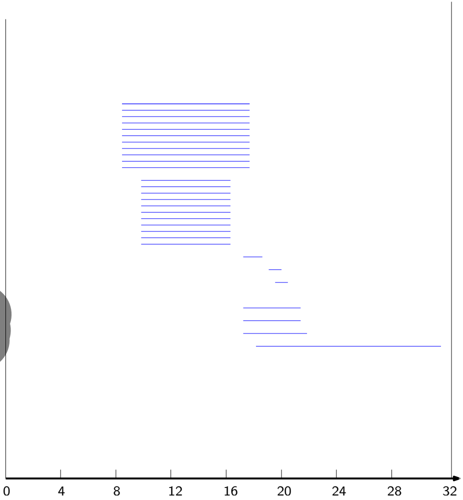
- less informative
- more stable
- generalizable



Dendrogram, Persistence Barcodes and Generalization

$$\begin{aligned} \mathrm{d}_{\mathcal{P}} : \quad & \mathbb{R}^2 \to \mathbb{R} \\ & x \mapsto \min_{p \in \mathcal{P}} \|x - p\| \end{aligned}$$





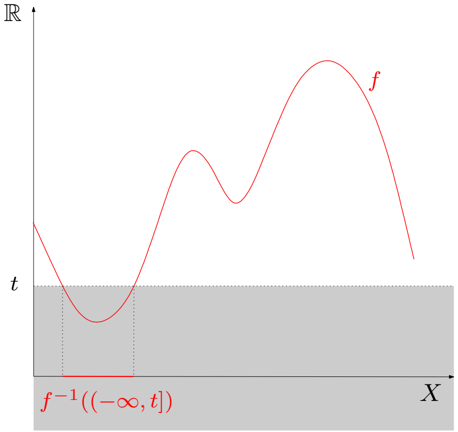
Barcode is:

- less informative
- more stable
- generalizable

Persistence Diagrams

Inside the black box:

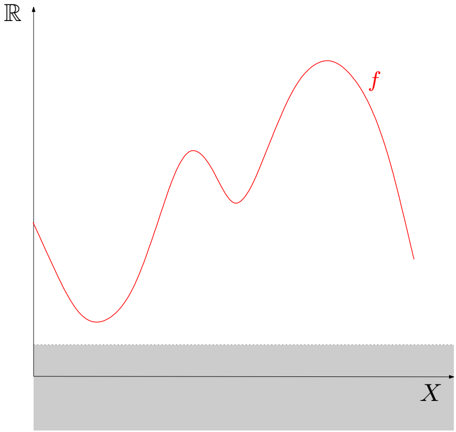
- Nested family (filtration) of sublevel-sets $f^{-1}\big((-\infty, t]\big)$, for $t \in \mathbb{R}$.
- Track the evolution of the topology (homology) of the family.



Persistence Diagrams

Inside the black box:

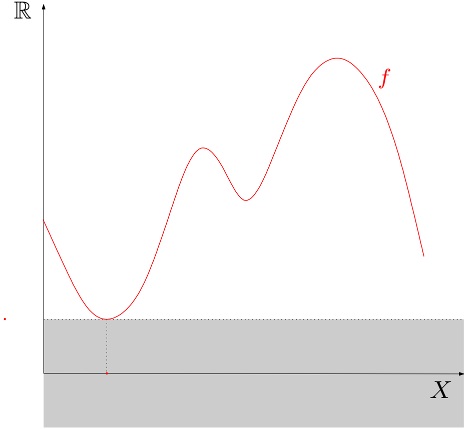
- Nested family (filtration) of sublevel-sets $f^{-1}\big((-\infty, t]\big)$, for $t \in \mathbb{R}$.
- Track the evolution of the topology (homology) of the family.



Persistence Diagrams

Inside the black box:

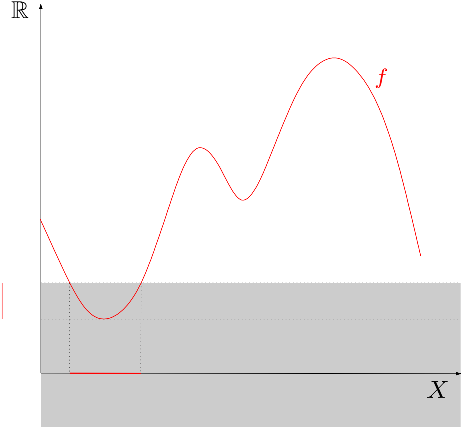
- Nested family (filtration) of sublevel-sets $f^{-1}\big((-\infty, t]\big)$, for $t \in \mathbb{R}$.
- Track the evolution of the topology (homology) of the family.



Persistence Diagrams

Inside the black box:

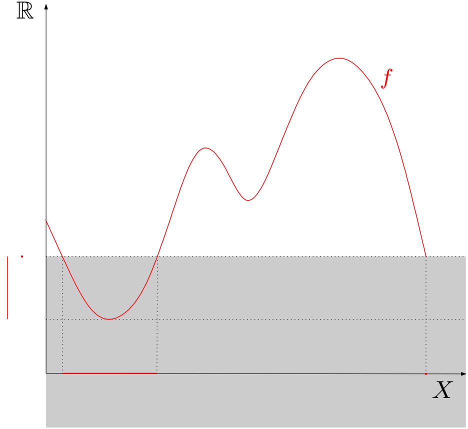
- Nested family (filtration) of sublevel-sets $f^{-1}\big((-\infty, t]\big)$, for $t \in \mathbb{R}$.
- Track the evolution of the topology (homology) of the family.



Persistence Diagrams

Inside the black box:

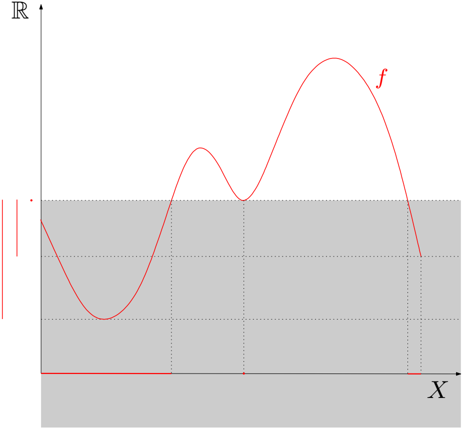
- Nested family (filtration) of sublevel-sets $f^{-1}\big((-\infty, t]\big)$, for $t \in \mathbb{R}$.
- Track the evolution of the topology (homology) of the family.



Persistence Diagrams

Inside the black box:

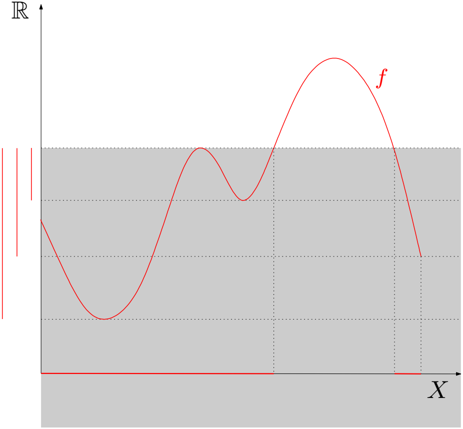
- Nested family (filtration) of sublevel-sets $f^{-1}\big((-\infty, t]\big)$, for $t \in \mathbb{R}$.
- Track the evolution of the topology (homology) of the family.



Persistence Diagrams

Inside the black box:

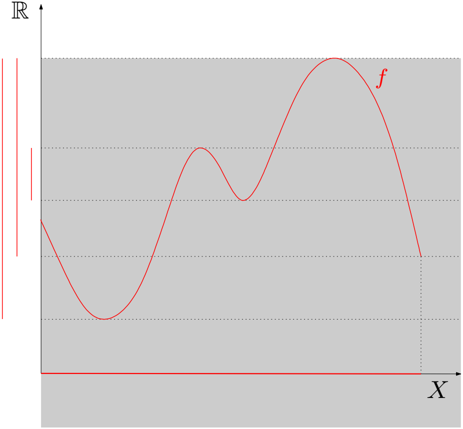
- Nested family (filtration) of sublevel-sets $f^{-1}\big((-\infty, t]\big)$, for $t \in \mathbb{R}$.
- Track the evolution of the topology (homology) of the family.



Persistence Diagrams

Inside the black box:

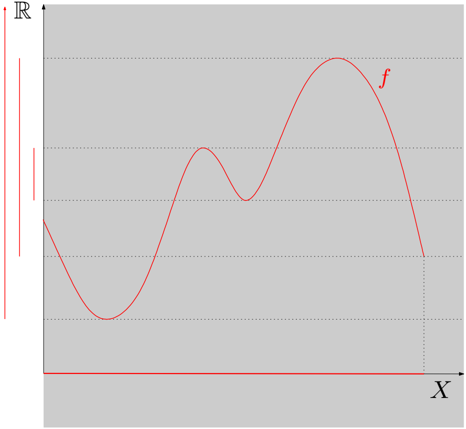
- Nested family (filtration) of sublevel-sets $f^{-1}\big((-\infty, t]\big)$, for $t \in \mathbb{R}$.
- Track the evolution of the topology (homology) of the family.



Persistence Diagrams

Inside the black box:

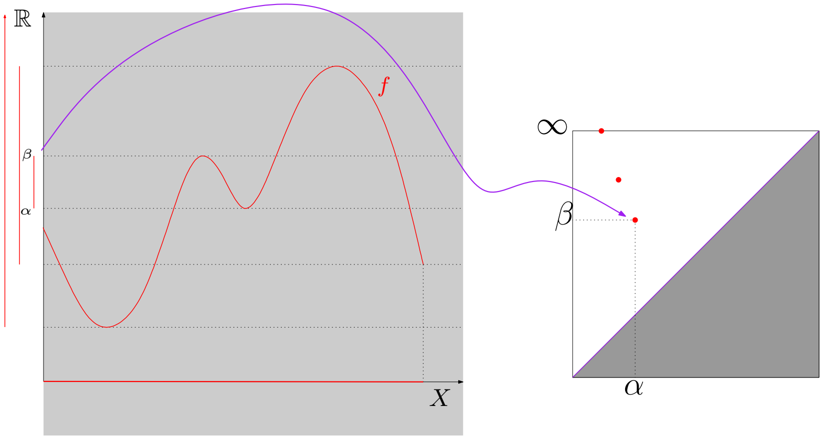
- Nested family (filtration) of sublevel-sets $f^{-1}\big((-\infty, t]\big)$, for $t \in \mathbb{R}$.
- Track the evolution of the topology (homology) of the family.



Persistence Diagrams

Inside the black box:

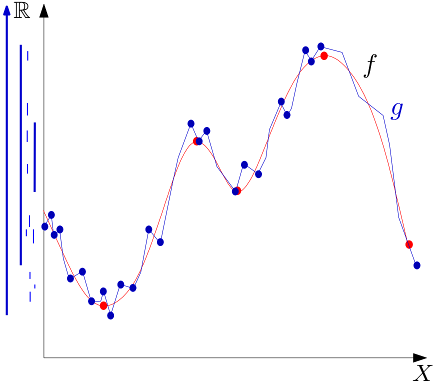
- Nested family (filtration) of sublevel-sets $f^{-1}\big((-\infty, t]\big)$, for $t \in \mathbb{R}$.
- Track the evolution of the topology (homology) of the family.

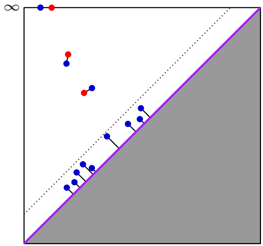


Persistence Diagrams

Inside the black box:

- Nested family (filtration) of sublevel-sets $f^{-1}\big((-\infty,t]\big)$, for $t \in \mathbb{R}$.
- Track the evolution of the topology (homology) of the family.



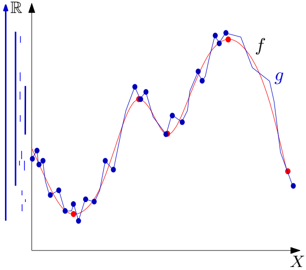


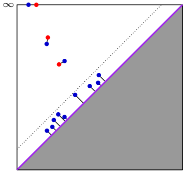
Persistence Diagrams

Definition (Bottleneck Distance)

Given two diagrams F and G,

$$\mathrm{d}_b(F,G) = \inf\{\delta \mid \textit{there exists a } \delta\textit{-correspondence between } F \textit{ and } G\}.$$



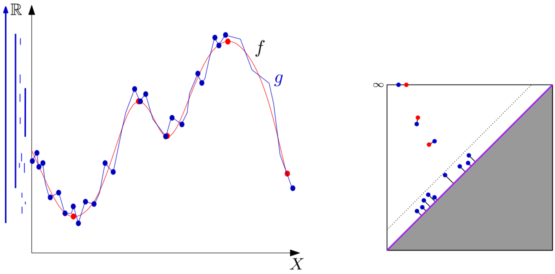


Persistence Diagrams

Definition (Bottleneck Distance)

Given two diagrams F and G,

$$\mathrm{d}_b(F, G) = \inf\{\delta \mid \textit{there exists a } \delta\textit{-correspondence between } F \textit{ and } G\}.$$



Theorem (Stability of Persistence)

For all nice functions $f, g : X \to \mathbb{R}$,

$$\mathrm{d_b}(\mathrm{dgm}(f), \mathrm{dgm}(g)) \leq \|f - g\|_\infty .$$

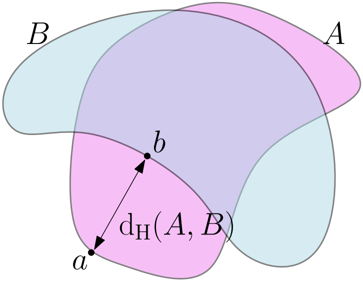
Stability for Sets

Definition (Hausdorff Distance)

The ***Hausdorff distance*** *between two compact sets* A *and* $B \subset \mathbb{R}^D$ *is*

$$\mathrm{d_H}(A, B) = \left\| \mathrm{d}_A(\cdot) - \mathrm{d}_B(\cdot) \right\|_\infty ,$$

where $\mathrm{d}_K(x) = \inf_{p \in K} \|x - p\|$ *is the distance to* K*.*



Stability for Sets

Definition (Hausdorff Distance)

*The **Hausdorff distance** between two compact sets* A *and* $B \subset \mathbb{R}^D$ *is*

$$\mathrm{d_H}(A, B) = \|\mathrm{d}_A(\cdot) - \mathrm{d}_B(\cdot)\|_\infty ,$$

where $\mathrm{d}_K(x) = \inf_{p \in K} \|x - p\|$ *is the distance to* K.

Proposition (Persistence Stability for Sets)

Write $\mathrm{dgm}(K)$ *for the diagram of the offset filtration*

$$K^r = \mathrm{d}_K^{-1}([0, r]), \text{ for } r \geq 0.$$

Then for all compact $A, B \subset \mathbb{R}^D$,

$$\mathrm{d_b}(\mathrm{dgm}(A), \mathrm{dgm}(B)) \leq \|\mathrm{d}_A(\cdot) - \mathrm{d}_B(\cdot)\|_\infty = \mathrm{d_H}(A, B).$$

Stability for Sets

Definition (Hausdorff Distance)

*The **Hausdorff distance** between two compact sets* A *and* $B \subset \mathbb{R}^D$ *is*

$$\mathrm{d_H}(A, B) = \|\mathrm{d}_A(\cdot) - \mathrm{d}_B(\cdot)\|_\infty ,$$

where $\mathrm{d}_K(x) = \inf_{p \in K} \|x - p\|$ *is the distance to* K*.*

Proposition (Persistence Stability for Sets)

Write $\mathrm{dgm}(K)$ *for the diagram of the offset filtration*

$$K^r = \mathrm{d}_K^{-1}([0, r]), \text{ for } r \geq 0.$$

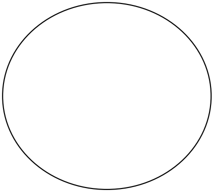
Then for all compact $A, B \subset \mathbb{R}^D$*,*

$$\mathrm{d_b}(\mathrm{dgm}(A), \mathrm{dgm}(B)) \leq \|\mathrm{d}_A(\cdot) - \mathrm{d}_B(\cdot)\|_\infty = \mathrm{d_H}(A, B).$$

Approximating persistence reduces to approximating sets for Hausdorff loss.

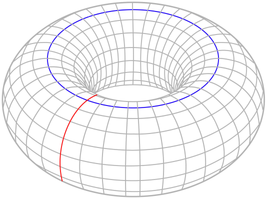
Homology in a Nutshell

β_0: connected components β_1: holes β_2: voids



$\beta_0 = 1$

$\beta_1 = 1$



$\beta_0 = 1$

$\beta_1 = 2$

$\beta_2 = 1$

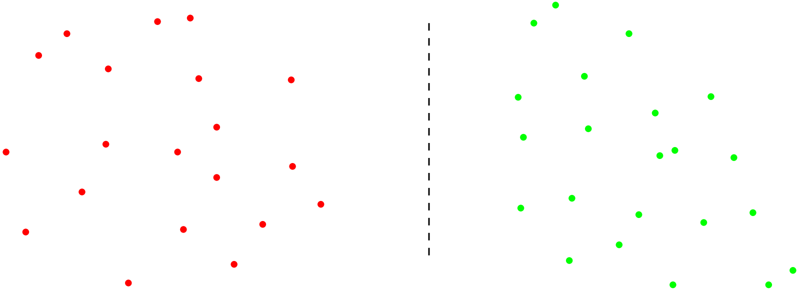
$\beta_0 = 1$

$\beta_1 = 1$

$\beta_2 = 0$

Support Estimation

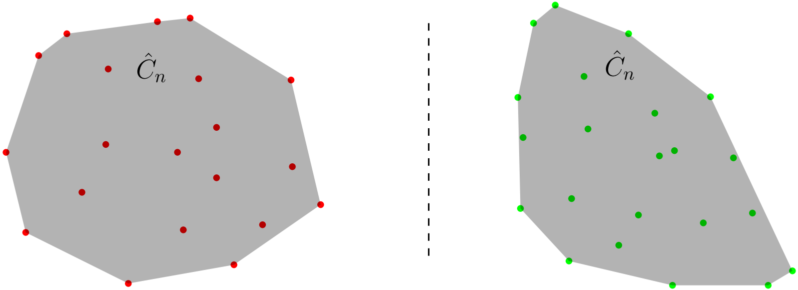
Data: A n-sample $X_1, \ldots, X_n \sim_{i.i.d.} P$.
Goal: Estimate the set $C = \text{Support}(P) = \bigcap_{\substack{K \subset \mathbb{R}^D \text{ closed} \\ P(K)=1}} K$.



Support Estimation

Data: A n-sample $X_1, \ldots, X_n \sim_{i.i.d.} P$.

Goal: Estimate the set $C = \text{Support}(P) = \bigcap_{\substack{K \subset \mathbb{R}^D \text{ closed} \\ P(K)=1}} K$.



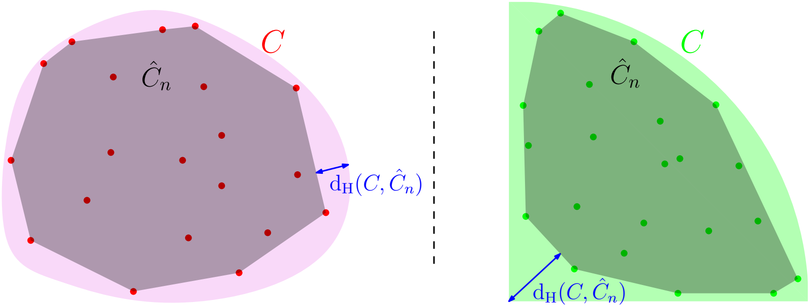
If we know (by advance) that C is convex, a good candidate is

$$\hat{C}_n = \text{Conv}(\{X_1, \ldots, X_n\}).$$

Support Estimation

Data: A n-sample $X_1, \ldots, X_n \sim_{i.i.d.} P$.

Goal: Estimate the set $C = \mathrm{Support}(P) = \bigcap_{\substack{K \subset \mathbb{R}^D \text{ closed} \\ P(K)=1}} K.$



If we know (by advance) that C is convex, a good candidate is

$$\hat{C}_n = \mathrm{Conv}(\{X_1, \ldots, X_n\}).$$

Support Estimation: Convex Case(s)

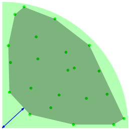
Theorem (Dümbgen, Walther – 1996)

Assume that $P = Unif_C$ is uniform over the convex set $C \subset \mathbb{R}^D$. Write

$$\hat{C}_n = \mathrm{Conv}(\{X_1, \ldots, X_n\}).$$

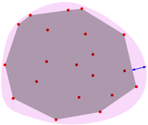
– *Then,*

$$\mathrm{d_H}(C, \mathbb{X}_n) \leq \mathrm{d_H}(C, \hat{C}_n) = O\left(\frac{\log n}{n}\right)^{\frac{1}{D}} \quad a.s.$$

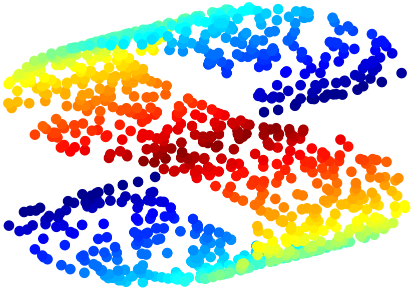


– *If in addition, ∂C is $\mathcal{C}^2$,*

$$\mathrm{d_H}(C, \hat{C}_n) = O\left(\frac{\log n}{n}\right)^{\frac{2}{D+1}} \quad a.s.$$



Beyond Convexity



How to model the support of these data?

- Low-dimensional and curved $\to$ Submanifold of $\mathbb{R}^D$.
- Not convex, but locally around it the projection uniquely defined.

Reminder: For a closed set $C \subset \mathbb{R}^D$,

$C \subset \mathbb{R}^D$ is convex $\Leftrightarrow$ Every $z \in \mathbb{R}^D$ has a unique nearest neighbor on C i.e. $\exists!\ \pi_C(z) \in C$ with $\|z - \pi_C(z)\| = \mathrm{d}_C(z)$.

Medial Axis

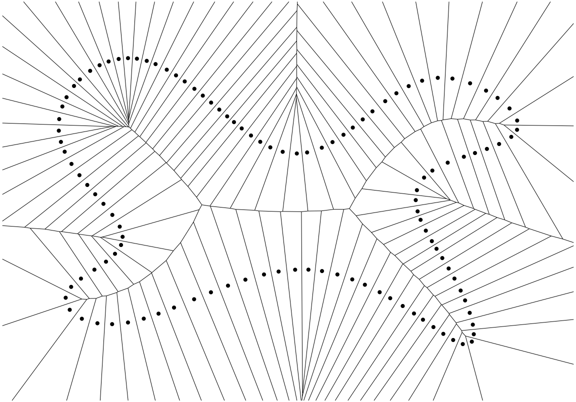
The **medial axis** of $M \subset \mathbb{R}^D$ is the set of points that have at least two nearest neighbors on M.

$$\mathrm{Med}(M) = \{z \in \mathbb{R}^D, z \text{ has several nearest neighbors on } M\},$$

Medial Axis

The **medial axis** of $M \subset \mathbb{R}^D$ is the set of points that have at least two nearest neighbors on M.

$$\mathrm{Med}(M) = \{z \in \mathbb{R}^D, z \text{ has several nearest neighbors on } M\},$$

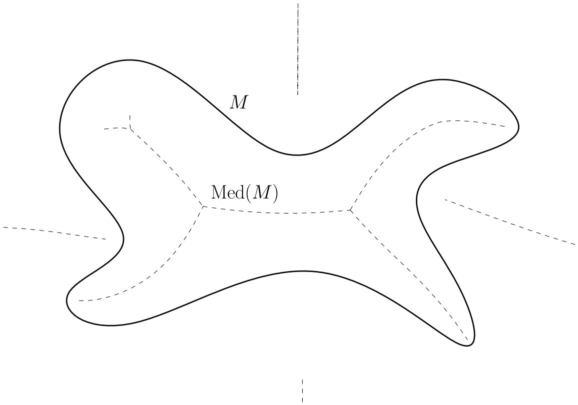


Medial axis of a point cloud (Voronoi faces)

Medial Axis

The **medial axis** of $M \subset \mathbb{R}^D$ is the set of points that have at least two nearest neighbors on M.

$$\mathrm{Med}(M) = \{z \in \mathbb{R}^D, z \text{ has several nearest neighbors on } M\},$$



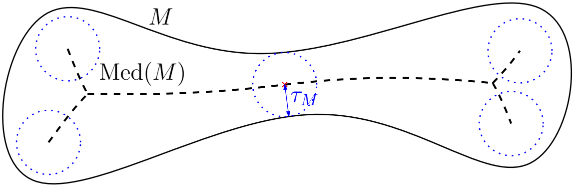
Medial axis of a continuous subset

Reach

For a closed subset $M \subset \mathbb{R}^D$, the **reach** τ_M of M is the least distance to its medial axis:

$$\tau_M = \inf_{x \in M} \mathrm{d}_{\mathrm{Med}(M)}(x),$$

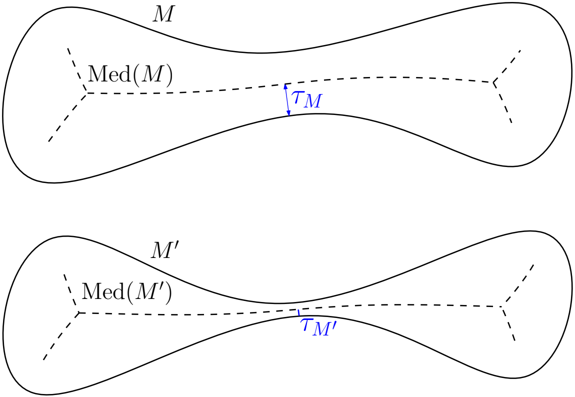
where for all $x \in \mathbb{R}^D$, $\mathrm{d}_K(x) = \inf_{p \in K} \|x - p\|$.



One can also flip the formula:

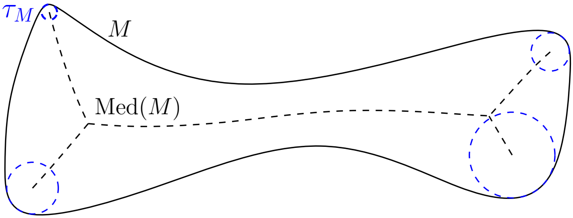
$$\tau_M = \inf_{z \in \mathrm{Med}(M)} \mathrm{d}_M(z).$$

Global Regularity



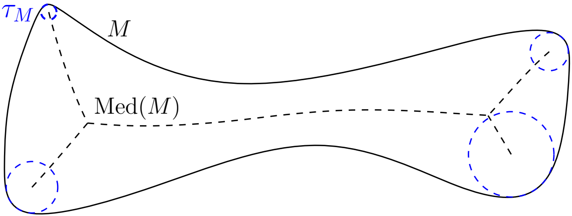
Narrow bottleneck structure $\Rightarrow \tau_M \ll 1$.

Local Regularity



High curvature $\Leftrightarrow$ Small radius of curvature $\Rightarrow \tau_M \ll 1$.

Local Regularity



High curvature $\Leftrightarrow$ Small radius of curvature $\Rightarrow \tau_M \ll 1$.

Proposition (Federer – 1959, Niyogi *et al.* – 2006)

Let II_x^M denote the second fundamental form of M.
For all unit tangent vector $v \in T_x M$, $\left\| II_x^M(v,v) \right\| \leq 1/\tau_M$.

As a consequence, the sectional curvatures κ of M satisfy

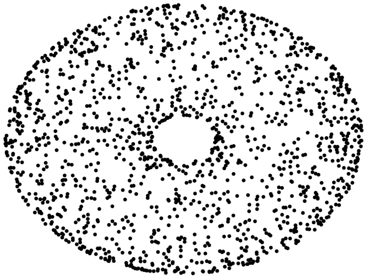
$$-2/\tau_M^2 \leq \kappa \leq 1/\tau_M^2.$$

Statistical Model

$X_1, \ldots, X_n \overset{i.i.d.}{\sim} P$, where $M = \mathrm{Support}(P) \subset \mathbb{R}^D$ satisfies:

- M is a compact connected d-dimensional submanifold,
- M has no boundary,
- $\tau_M \geq \tau_{\min} > 0$,
- P is (almost) the uniform distribution on M.

The set of distributions satisfying these conditions is denoted by $\mathcal{P}$.



A Reconstruction Theorem

Theorem (A, Levrard – 2018)

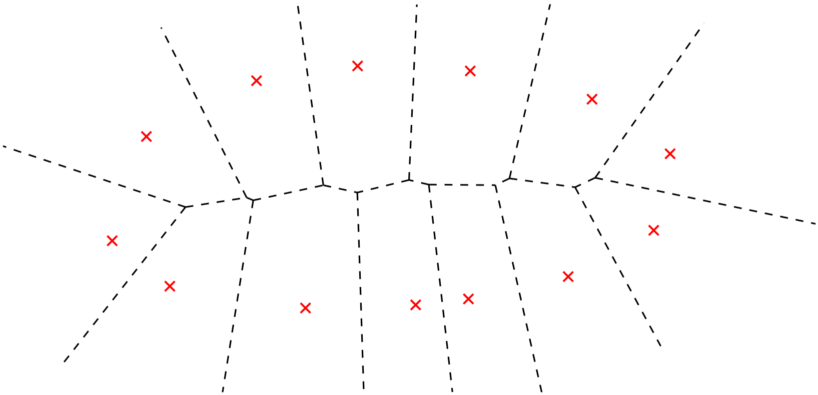
There exists a computable estimator $\hat{M}$ such that for all $n \geq 1$,

$$\mathbb{E}_{P^n}\left[\mathrm{d_H}(M, \hat{M})\right] \leq C \left(\frac{\log n}{n}\right)^{2/d},$$

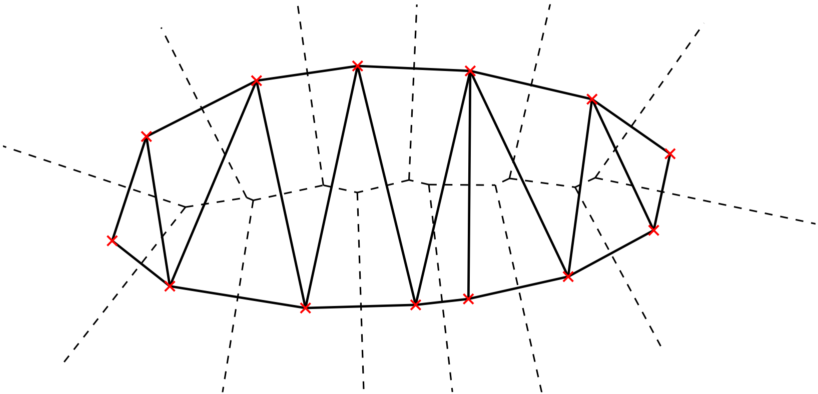
where $C = C_{\tau_{\min},d}$ does not depend on the ambient dimension D.

Outline of the Method

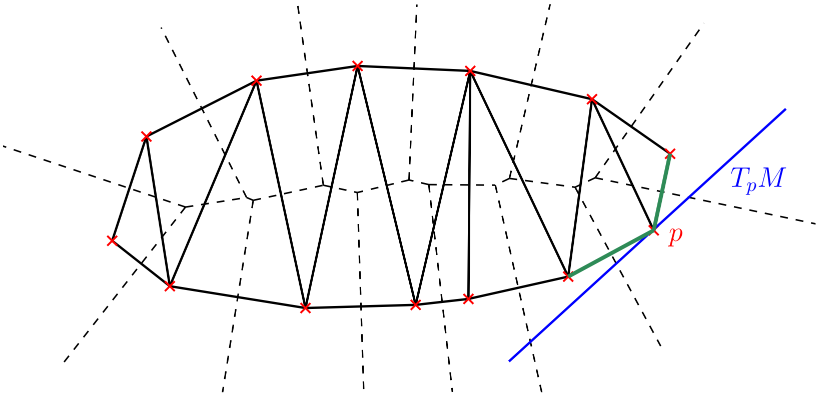
Outline of the Method



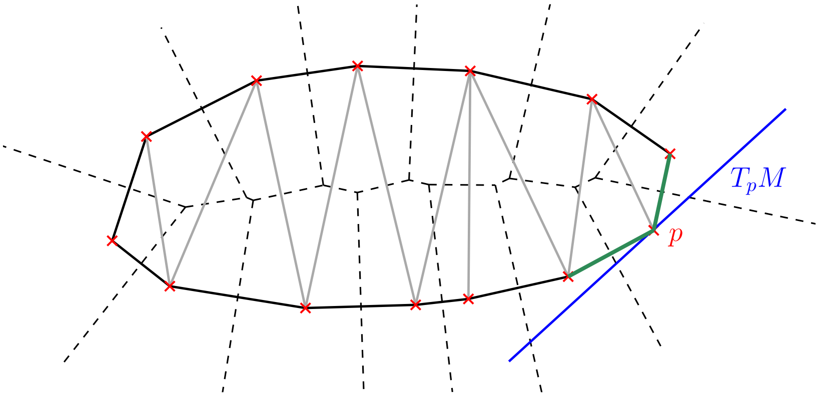
Outline of the Method



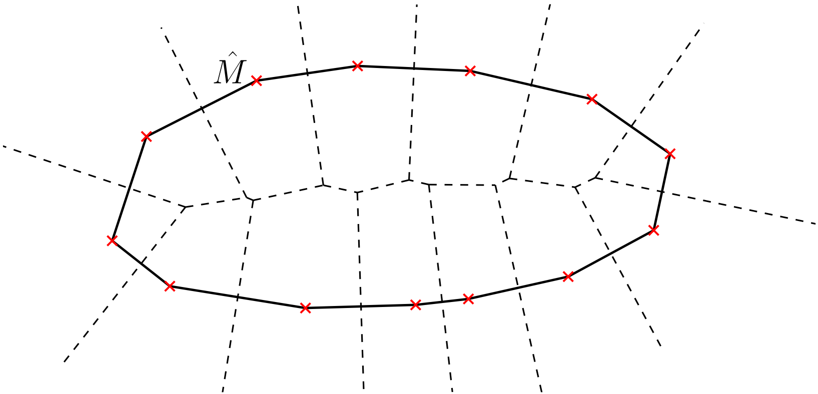
Outline of the Method



Outline of the Method



Outline of the Method



The Tangential Delaunay Complex [Boissonnat & Ghosh – 2014]

Optimality: Studying the Minimax Risk

The **minimax risk** over the statistical model $\mathcal{P}$ is

$$\inf_{\hat{M}_n} \sup_{P \in \mathcal{P}} \mathbb{E}_{P^n} \left[\mathrm{d_H}\big(M, \hat{M}_n\big) \right],$$

where the infimum is taken over all the estimators $\hat{M}_n = \hat{M}_n(\mathbb{X}_n)$ computed over a n-sample $\mathbb{X}_n = \{X_1, \ldots, X_n\}$.

Optimality: Studying the Minimax Risk

The **minimax risk** over the statistical model $\mathcal{P}$ is

$$\inf_{\hat{M}_n} \sup_{P \in \mathcal{P}} \mathbb{E}_{P^n}\left[\mathrm{d_H}\big(M, \hat{M}_n\big)\right],$$

where the infimum is taken over all the estimators $\hat{M}_n = \hat{M}_n(\mathbb{X}_n)$ computed over a n-sample $\mathbb{X}_n = \{X_1, \ldots, X_n\}$.

Proposition (Genovese *et al* – 2012)

For n large enough,

$$\inf_{\hat{M}_n} \sup_{P \in \mathcal{P}} \mathbb{E}_{P^n}\left[\mathrm{d_H}\big(M, \hat{M}_n\big)\right] \leq C \left(\frac{\log n}{n}\right)^{\frac{2}{d}},$$

where $C = C_{d,\tau_{\min}}$

Optimality: Studying the Minimax Risk

The **minimax risk** over the statistical model $\mathcal{P}$ is

$$\inf_{\hat{M}_n} \sup_{P \in \mathcal{P}} \mathbb{E}_{P^n} \left[\mathrm{d_H}\big(M, \hat{M}_n\big) \right],$$

where the infimum is taken over all the estimators $\hat{M}_n = \hat{M}_n(\mathbb{X}_n)$ computed over a n-sample $\mathbb{X}_n = \{X_1, \ldots, X_n\}$.

Proposition (Genovese *et al* – 2012)

For n large enough, (+ mild technical assumptions)

$$c \left(\frac{1}{n}\right)^{\frac{2}{d}} \leq \inf_{\hat{M}_n} \sup_{P \in \mathcal{P}} \mathbb{E}_{P^n} \left[\mathrm{d_H}\big(M, \hat{M}_n\big) \right] \leq C \left(\frac{\log n}{n}\right)^{\frac{2}{d}},$$

where $C = C_{d,\tau_{\min}}$ and $c = c_{\tau_{\min}}$.

Lower Bound Technique: Le Cam's Lemma

Theorem (L. Le Cam)

For all $P_0, P_1 \in \mathcal{P}$,

$$\inf_{\hat{M}_n} \sup_{P \in \mathcal{P}} \mathbb{E}_{P^n}\left[\mathrm{d_H}\big(M, \hat{M}_n\big)\right] \geq \frac{1}{2}\mathrm{d_H}(M_0, M_1)\big(1 - \mathrm{TV}(P_0, P_1)\big)^n,$$

where

$$\mathrm{TV}(P_0, P_1) = \sup_{B \in \mathcal{B}(\mathbb{R}^D)} |P_0(B) - P_1(B)|$$

denotes the total variation distance between P_0 *and* P_1.

Lower Bound Technique: Le Cam's Lemma

Theorem (L. Le Cam)

For all $P_0, P_1 \in \mathcal{P}$,

$$\inf_{\hat{M}_n} \sup_{P \in \mathcal{P}} \mathbb{E}_{P^n}\left[\mathrm{d_H}\big(M, \hat{M}_n\big)\right] \geq \frac{1}{2}\mathrm{d_H}(M_0, M_1)\big(1 - \mathrm{TV}(P_0, P_1)\big)^n,$$

where

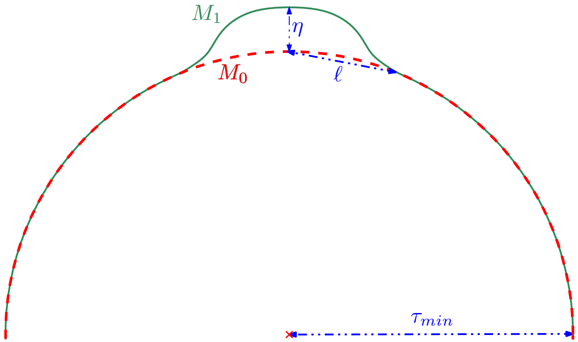
$$\mathrm{TV}(P_0, P_1) = \sup_{B \in \mathcal{B}(\mathbb{R}^D)} |P_0(B) - P_1(B)|$$

denotes the total variation distance between P_0 *and* P_1*.*

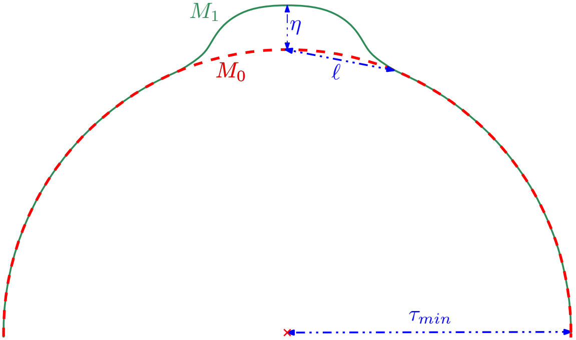
Deriving a good lower bound amounts to find P_0, P_1 such that:

- $P_0, P_1 \in \mathcal{P}$,
- $\mathrm{d_H}(M_0, M_1)$ is large,
- $\mathrm{TV}(P_0, P_1)$ is small.

Le Cam's Lemma Heuristic

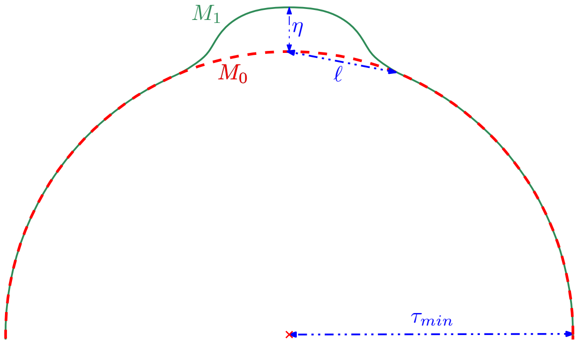


Le Cam's Lemma Heuristic



- P_0 and P_1 both belong to $\mathcal{P}$ as soon as $\eta \lesssim \ell^2$,
- $\mathrm{d_H}(M_0, M_1) \geq \eta$,
- $\mathrm{TV}(P_0, P_1) \lesssim \ell^d$.

Le Cam's Lemma Heuristic

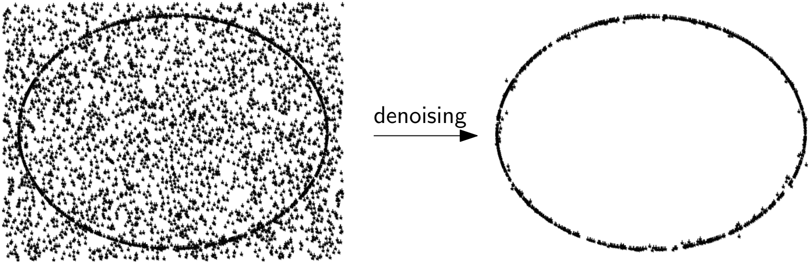


- P_0 and P_1 both belong to $\mathcal{P}$ as soon as $\eta \lesssim \ell^2$,
- $\mathrm{d_H}(M_0, M_1) \geq \eta$,
- $\mathrm{TV}(P_0, P_1) \lesssim \ell^d$.

Hence, for $\eta \approx \ell^2$ and $\ell \approx (1/n)^{1/d}$,

$$\inf_{\hat{\tau}_n} \sup_{P \in \mathcal{P}} \mathbb{E}_{P^n}\left[\mathrm{d_H}\left(M, \hat{M}_n\right)\right] \gtrsim \eta \left(1 - \ell^d\right)^n \approx (1/n)^{2/d}.$$

Extension to a Noisy Model

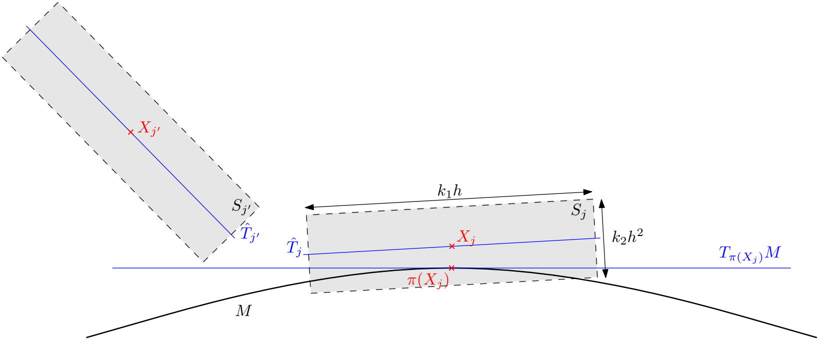


Theorem (A, Levrard – 2018)

For all $\delta > 0$, there exists a computable estimator $\hat{M}_n^{(\delta)}$ such that for all $n \geq 1$,

$$\mathbb{E}\big[\mathrm{d_H}(M, \hat{M}_n^{(\delta)})\big] \leq C \left(\frac{\log n}{n}\right)^{2/d-\delta}.$$

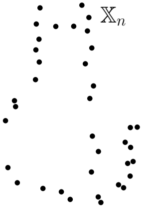
Denoising Outline



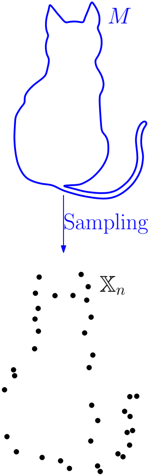
$$\begin{aligned} P\left(S(x, T_{\pi(x)}M)\right) &\asymp h^d & &\text{if} \quad \mathrm{d}(x,M) \leq h^2, \\ P\left(S(x,T)\right) &\asymp h^{2D-d} & \text{for all } T, \ &\text{if} \quad \mathrm{d}(x,M) > h^2, \end{aligned}$$

Since $h^{2D-d} \ll h^d$, the measure $P\left(S(x,T)\right)$ of the slabs are discriminative for denoising.

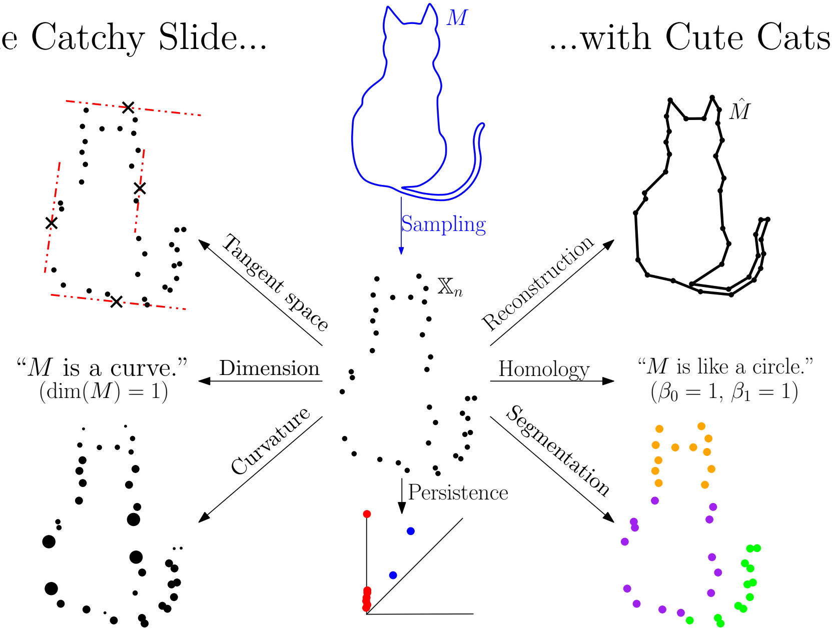
The Catchy Slide...



The Catchy Slide... ...with Cute Cats



The Catchy Slide... ...with Cute Cats



The Catchy Slide... ... with Buzzwords

Lots of theoretical related topics:

- High-Dimensional statistics
- Nonparametric statistics
- Time series
- Computational geometry
- Geometry processing
- Abstract algebra

With applications in

- Material science
- Image analysis
- Physical chemistry
- Cosmology
- Network analysis
- ...